

Does a Computer-Based System Help Researchers in Medical Genetics to Perform Familial Studies Without Manipulating Nominal Data?

Steve Arsenault, B.Sc., Daniel Gaudet, M.D

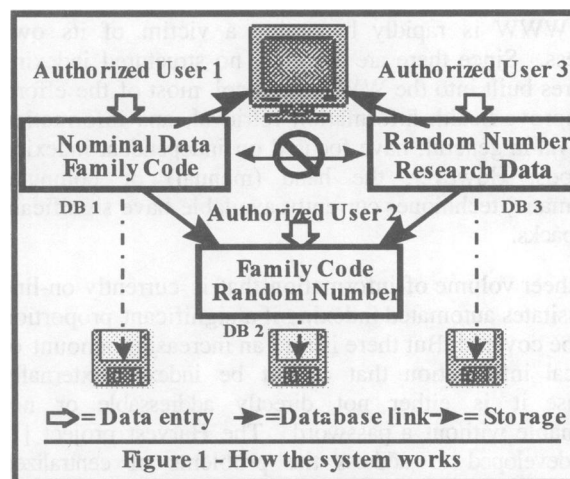
Chicoutimi Hospital Research Unit, Chicoutimi, Canada

Background. Numerous strategies have been developed over the years for the codification of nominal medical data (1, 2). However, patient follow-up in the medical genetic's study field, raises a problem in the linking and use of nominal familial informations. Nevertheless, the confidentiality of individual as well as of familial data must be respected. We have created a system with which familial studies can be performed without manipulating nominative data.

System description. The object of this system consists in preventing disclosure and preserving integrity of familial data. The proposed system systematically codifies familial links and generates 3 different databases simultaneously, in order to ensure the confidentiality of research data (figure 1). In the first database (DB1), all nominative informations are matched with a family code. The second database (DB2) includes the family code and an individual random number identifying each individual in the research project. The third one (DB3) links the random number to research data.

The 3 generated databases are stored in independant mediums, increasing the level of security (referential integrity) and confidentiality (prevention of disclosure). Authorized users can create virtual links between any combination of 2 or 3 generated databases in order to add new data, check the identity of individuals or validate the familial links. However, the command required to link the 3 databases simultaneously is under the control of 3 keys: the first key to be defined by the treating physician (access to DB1), the second one by the principal investigator (access to DB2) and the third one by the research coordinator (access to DB3). A new patient name and the identification of a known subject already in the database are the only required entries needed by the system to operate. Depending on the nature of the research (clinical vs epidemiological), nominal informations in the DB1 can be anonymized (irreversible encryption).

The system has been developed using Microsoft Visual Basic 4.0 and structured query language (SQL) statements. The system uses simple WYSIWYG graphical interface. The family codes are compatible with popular softwares generating family trees such as Cyrillic 2.0.



Evaluation. In 1995 and 1996, this codification strategy was tested in 9 research projects involving 1968 subjects from 560 families. On average, each project involved 260 subjects. The largest family included 127 individuals. From 1995 to 1996, the number of families for which the codification has been used increased by 30%.

Conclusion. This system is designed to preserve individual confidentiality as well as familial data integrity and security. The physical separation of nominal and research data provides an effective protection against hacking and cracking. In this regard, the use of this computer-based system for the realization of familial studies without nominative data manipulation might be helpful for researchers in medical genetics.

References

1. Quantin C, Bouzelat H, Dusserre L. Irreversible encryption method by generation of polynomials. *Med inform* 1996;21:113-121.
2. Michaelis J, Miller M, Pommerening K and Schmidtman I. A new concept to ensure data privacy and data security in cancer registries. *InMEDINFO 95: Proceedings of the eighth World Congress on Medical Informatics* Greenes RA, Peterson HE and Protti DJ (eds) (North-Holland: Elsevier) pp 661-665.